

APPLICATION FOR UNITED STATES LETTERS PATENT

FOR

**CENTRALIZED CELL HOMING AND LOAD BALANCING IN A BASE  
STATION CONTROLLER**

Inventors: Mun-Choon Chan  
Martin Havemann  
Ramachandran Ramjee  
Sandra R. Thuel  
Thomas Y. Woo

Prepared by: Mendelsohn & Associates, P.C.  
1515 Market Street, Suite 715  
Philadelphia, Pennsylvania 19102  
(215) 557-6656

\* \* \* \* \*

Certification Under 37 CFR 1.10

"Express Mail" Mailing Label No. EV140153387US

Date of Deposit March 31, 2004.

I hereby certify that this document is being deposited with the United States Postal Service's "Express Mail Post Office To Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313.

Mary E. Caniz  
(Name of person mailing)

Mary E. Caniz  
(Signature of person mailing)

## **CENTRALIZED CELL HOMING AND LOAD BALANCING IN A BASE STATION CONTROLLER**

### 5 BACKGROUND OF THE INVENTION

#### Field of the Invention

The present invention relates to allocation of processing in a telecommunications network, and, more particularly, to allocation of connections to processors in a base station controller of a wireless network.

#### 10 Description of the Related Art

Radio or wireless telecommunications networks provide wireless voice and data communication between different service providers, such as the public switched telephone network (PSTN) and the Internet. Typical networks deployed in the prior art operate in accordance with General Packet Radio Service (GPRS), Universal Mobile  
15 Telecommunications Systems (UMTS), and Code-Division Multiple Access (CDMA) 2000 network standards. A wireless network is typically characterized by one or more base stations (BSs), a radio network controller (RNC), a mobile switching center (MSC), and data gateway. Each BS manages at least one corresponding cell, or dedicated area of wireless radio coverage. The MSC serves as an interface to, for example, the PSTN  
20 network, while a data gateway serves as an interface to, for example, an IP network that connects to the Internet. Each base station communicates with user elements (UEs) through wireless radio signals. A user element is either a fixed or mobile transceiver, such as a mobile phone, wireless notebook computer, or fixed radio station.

The RNC includes a base station controller (BSC) and a traffic-processing unit  
25 (TPU). The TPU is employed for management of voice and data forwarding (traffic), while the BSC is employed for management of call control (signaling). The BSC connects the CDMA radio access network (RAN) with the core network (CN). Thus, the BSC manages i) the radio path and ii) signaling between each UE and the CN.

The BSC comprises multiple central processing units (CPUs) distributed over  
30 multiple blades (a blade is circuitry of predetermined functionality assigned to cells or cell sectors, such as equipment processor cards or units). Each CPU runs a call-processing

application (software program) that, besides tasks such as soft-handover (mobility), outer-loop power control, measurement reports, and radio channel reconfiguration, guides the process of setting up and tearing down a user call. The call-processing application processes call-setup and call-release signaling messages, and the call-processing

5 application also manages cell components. A cell component is a logical representation of circuitry and related processing that enables per sector per carrier coverage of a BS cell area (cells may be divided into sectors). For example, in CDMA networks, cell components manage spreading code trees for spreading code assignments to a cell/sector. Cell components employ downlink spreading codes used by BSs to set up downlink

10 channels to UEs (e.g., mobile phones). In a soft-handover/soft-handoff situation, multiple BSs are connected simultaneously to the UE, where these BSs use spreading codes with the same spreading factor when communicating with this UE. Cell components of blades communicate with one another or are polled by a coordinating instance in order to determine free codes (i.e., available and unused) with the same spreading factor..

15 CPUs of a prior art BSC are associated with specific cells and call processing applications. Calls originating in a cell are generally handled by a call processing application running on the same CPU. Depending on the cell homing strategy (the configuration method of how to assign cell components to CPUs) of the BSC and user mobility patterns (the way users move through cells during active calls), some CPUs of

20 the prior art BSC might become highly loaded, while other CPUs might be underutilized if a particular cell has a large number of voice and data calls. Cell homing strategies could be, but are not limited to, a random distribution so that all CPUs manage the same amount of Cell components, or a load biased distribution based on historical data so that heaviest loaded cells are not associated with the same CPU. The distributed cell homing strategy

25 requires repeated traffic-load analysis to define an optimal cell placement because of changes in user mobility patterns, charging models, and new mobile applications.

Cell components generate CPU load whenever spreading codes are allocated or released. In a distributed system, the call processing application sends a code allocation/release request individually to all cell components holding a soft-handover leg

30 of the processed UE. A voice call sets up a traffic connection right after the signaling connection is established. Both connections are torn down at voice call-release time. A

data call also sets up both connections one after each other. However, after data session initialization, the data call immediately suspends the traffic connection if there is no data to be sent. This suspension state means that spreading codes are released, which is transparent to the data exchanging application on the UE. Only when a burst of data is to be sent new spreading codes are allocated and the traffic radio link is revived. When the data transmission is complete, the traffic connection (radio link) is again suspended. The bursty nature of data traffic causes a multitude of radio link setups and teardowns, which in turn triggers spreading code allocations and releases.

In addition, when a UE crosses a cell boundary, a new CPU, potentially located on a different blade, may serve the radio portion of the connection. However, the call-processing application CPU association, termed UE context, of the original CPU that served the UE's call-set-up request does not move except for exceptional conditions like switching the anchor/serving RNC, because it would be computationally very expensive to move all the call state information. Since the UE context does not move, messages are generated between the original and new CPUs, causing higher inter-CPU messaging overhead.

### SUMMARY OF THE INVENTION

In accordance with exemplary embodiments of the present invention, a base station controller (BSC) of a radio or wireless telecommunications network includes a director. A BSC includes multiple central processing units (CPUs), with each CPU running a call-processing application for one or more connections. The director intercepts wireless call-setup signaling and assigns each corresponding connection to a CPU according to a centralized load-balancing algorithm. The centralized load-balancing algorithm distributes connections to less-loaded CPUs to i) prevent individual CPUs from overloading, ii) utilize otherwise unused system resources, and iii) increase overall system performance. The director hosts cell components that manage code division multiple access (CDMA) downlink spreading codes for a base station, providing centralized allocation of spreading codes for the call processing application. The call processing application sends a code allocation request to the director, which returns a set of codes for base stations currently connected to the UE.

In accordance with exemplary embodiments of the present invention, processing capacity of processors in a radio network controller is allocated by: 1) monitoring for a message of a connection between a user element and a network and 2) allocating, if the message is a call set-up message, one of the processors to the connection in accordance with a load-balancing algorithm, and, if the message is an allocation message, a set of spreading codes to the connection with the same spreading factor.

### BRIEF DESCRIPTION OF THE DRAWINGS

Other aspects, features, and advantages of the present invention will become more fully apparent from the following detailed description, the appended claims, and the accompanying drawings in which:

FIG. 1 shows a block diagram of communications network with centralized cell homing and load balancing in accordance with exemplary embodiments of the present invention;

FIG. 2 shows an exemplary embodiment of the radio network controller of FIG 1; and

FIG. 3 shows an exemplary method employed by the director of FIG. 2.

### DETAILED DESCRIPTION

FIG. 1 shows a block diagram of communications network 101 with centralized cell homing and load balancing in accordance with exemplary embodiments of the present invention. Network 101 might be employed to provide wireless voice and data communication between different service providers, shown as public switched telephone network (PSTN) 102 and Internet 103. Network 101 might operate in accordance with General Packet Radio Service (GPRS), Universal Mobile Telecommunications Systems (UMTS) network, and/or Code-Division, Multiple Access (CDMA) 2000 standards. Network 101 comprises at least two base stations (BSs) 110 and 111, radio network controller (RNC) 112, mobile switching center (MSC) 113, and data gateway 114.

RNC 112 includes a director operating in accordance with an exemplary embodiment of the present invention, as described subsequently. BS 110 provides

wireless radio communication for cells 120 and 121, BS 111 provides wireless radio communication for cells 122 and 123, and data gateway 114 may comprise serving support node 115 and gateway support node 116. The data gateway manages user mobility, billing record collection, IP address allocation, etc. for the data calls. Each base station may communicate with user element (UE) 130 through wireless radio links 131 and 132.

FIG. 2 shows an exemplary embodiment of RNC 112 having base station controller (BSC) 201, traffic processing unit (TPU) 202, and switch 203. BSC 201 includes director 210 operating in accordance with one or more exemplary embodiments of the present invention. TPU 202 is employed for management of voice and data forwarding, while BSC 201 is employed for management of call control (signaling). Switch 203 is a low-level component employed for connectivity to transfer signals between BSC 201 and TPU 202.

BSC 201 connects the CDMA radio access network (RAN) with the core network (CN). In FIG. 1, the RAN includes UE 130, BS 110, and BS 111, and the CN includes MSC 113 coupled to PSTN 102 and data gateway 114 coupled to an IP network 103. A provider's internal IP network typically connects the data gateway to the Internet through a firewall. BSC 201 coordinates operation of the radio paths and signaling between UE 130 and the CN.

Returning to FIG. 2, BSC 201 comprises multiple central processing units (CPUs) 203(1) through 203(N) distributed over multiple blades 204(1) through 204(M). Each of CPUs 203(1) through 203(N) runs an instance of a call-processing application that guides the process of setting up and tearing down a UE connection (or "call", though the connection may be voice, voiceband data, or packet data) with the CN.

A call-processing application running on a given one of CPUs 203(1) through 203(N) processes call-setup and call-release signaling messages. The call-processing application also manages cell components. A cell component enables per sector per carrier coverage of a BS cell area. Cell components employ downlink spreading codes used by BSs to set up downlink channels to UEs (e.g., mobile phones). In soft-handover/soft-handoff multiple BSs are connected simultaneously to the UE using spreading codes with the same spreading factor. Cell components communicate with

director 210, which determines free codes with the same spreading factor.

In accordance with exemplary embodiments of the present invention, director 210 is a logical entity representing circuitry and related processing that monitors messages that are employed to setup a connection or relate to soft-handover/soft-handoff of the connection between cells/sectors. Director 210 intercepts wireless call-setup signaling and assigns the corresponding connection to one of the CPUs 203(1) through 203(N) according to a centralized load-balancing algorithm. The centralized load-balancing algorithm distributes connections to less-loaded CPUs to i) prevent individual CPUs from moving to an overload state, ii) utilize otherwise unused system resources, and iii) increase overall system performance. Once the call is assigned to a CPU, director 210 forwards all subsequent call related signaling messages to the CPU for processing by a call-processing application.

Director 210 also serves as a host to cell components that manage CDMA downlink spreading codes for BS 110. Director 210 provides centralized allocation of spreading codes for the call processing application running on CPU 203. Director 210 responds to allocation/release messages for spreading codes to establish/teardown radio links (e.g. soft-handover/soft-handoff, data bursts, voice call setup/release) and queries the corresponding cell components to determine an appropriate set of spreading codes with the same spreading factor. Once the set of spreading codes for soft-handover/soft-handoff of the connection is determined, director 210 returns the set of spreading codes to the corresponding querying call processing application running on CPU 203.

While director 210 is shown in FIG. 2 as a separate component of BSC 201, the present invention is not so limited. The director might be embodied in any of a number of components within the RNC, such as within a dedicated processor on a given blade, or as a function implemented in any one of the processors within the RNC that is coupled to the BSC.

FIG. 3 shows an exemplary method 300 employed by director 210 of FIG. 2. At step 301, the director receives a call-setup request or an allocation message. At step 302, a test determines whether a call-setup request or an allocation message was received. If the test of test 302 determines that a call-setup request was received, then the method advances to step 304. If the test of test 302 determines that an allocation message was

received, then the method advances to step 306.

At step 304, the director selects a BSC CPU for call processing based on a predefined load-balancing algorithm. At step 305, the director forwards the call-setup request to the call-processing application running on the selected CPU. The  
5 call-processing application may then process the call-setup request. From step 305, the method returns to step 301.

At step 306, the director processes the allocation message. For downlink spreading code allocation, the director typically receives a single allocation message. The director queries the cell components to determine a set of spreading codes with the same  
10 spreading factor. The set of spreading codes depends on the number of soft-handover/soft-handoff legs. At step 307, the director sends the set of spreading codes with the same spreading factor to the call-processing application. From step 307, the method returns to step 301.

The director at step 304 might employ different load-balancing algorithms known  
15 in the art. For example, two load-balancing algorithms are as follows. The first load-balancing algorithm allocates based on the call-context amount per CPU, where the law of large numbers results in an averaging of the load imposed by individual users. This first load-balancing algorithm keeps track how many calls are currently processed on each blade (e.g., either by intercepting call-setup/call-release messages or by periodic polling of  
20 all CPUs/blades). The algorithm determines the average number of calls per CPU (= total calls / CPUs), and this average number is weighted by the total call capacity of the CPU. The CPU with the least weighted calls is then selected. Because the number of calls in the system is high (several thousand), the system load imposed by the calls average out.

The second load-balancing algorithm balances based on CPU utilization weighted  
25 with the call processing capacity. Beyond a certain threshold of CPU utilization, a connection is redirected to the least-loaded CPU. Redirection continues until the CPU utilization sinks below the threshold or a close to overload threshold is reached. At this second threshold redirection is stopped and new calls begin to be dropped. Load-balancing algorithms might require a measure of CPU loading as an indicator of CPU  
30 utilization. Such measure might be, for example, a measure of the number of calls assigned to a CPU, a measure of the type (e.g., voice, data, or voiceband data) of

connections processed by a CPU, or a measure of the processor utilization (amount of available processing capacity or “power”) currently used by the CPU.

Centralized load balancing and cell homing in a BSC in accordance with one or more embodiments of the present invention may provide for the following advantages.

- 5 Centralized information about system load enables use of optimized load-balancing algorithms. Such load-balancing algorithm might be optimized only for CPU utilization without considering varying message costs to cell components caused by a distributed cell homing strategy. For a connection in a soft-handover/soft-handoff state, negotiation of assigned spreading codes might be performed locally, which reduces inter-CPU overhead
- 10 message traffic. Also, the present invention allows for lower operational costs since no historical traffic data analysis is required for connection and spreading-code assignments, reducing the complexity associated with adding new cells.

- The present invention may be embodied in a processor, such as a network controller or computer, and the processor may be coupled to a network or network
- 15 database to receive network topology, provisioning, and capacity information used by the methods as described herein. In addition, the present invention may be employed either for wired, wireless, optical, or non-optical networks, and may be employed for either synchronous or asynchronous networks.

- As would be apparent to one skilled in the art, the various functions of the director
- 20 may be implemented with circuit elements or may also be implemented in the digital domain as processing steps in a software program. Such software may be employed in, for example, a digital signal processor, micro-controller, or general-purpose computer.

- The present invention can be embodied in the form of methods and apparatuses for practicing those methods. The present invention can also be embodied in the form of
- 25 program code embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. The present invention can also be embodied in the form of program code, for example, whether stored in a storage medium, loaded into
- 30 and/or executed by a machine, or transmitted over some transmission medium, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation,

wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. When implemented on a general-purpose processor, the program code segments combine with the processor to provide a unique device that operates analogously to specific logic  
5 circuits.

It will be further understood that various changes in the details, materials, and arrangements of the parts which have been described and illustrated in order to explain the nature of this invention may be made by those skilled in the art without departing from the principle and scope of the invention as expressed in the following claims.